Probabilistic Inference Models

ITI0210, lecture 10 (2021)

Review

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	.108	.012	.072	.008
¬ *cavity*	.016	.064	.144	.576

Give evidence $\boldsymbol{e}$, we can infer

the probability distribution of query variables $\boldsymbol{q}$

$$P(\boldsymbol{q}|\boldsymbol{e}) = \frac{P(\boldsymbol{q},\boldsymbol{e})}{P(\boldsymbol{e})} = \frac{\sum_{\boldsymbol{h}'} P(\boldsymbol{q},\boldsymbol{e},\boldsymbol{h}')}{\sum_{\boldsymbol{q}',\boldsymbol{h}'} P(\boldsymbol{q}',\boldsymbol{e},\boldsymbol{h}')}$$

Computational Issues

1. Joint probability distribution is exponential in size
   (example: 2^n for n Boolean variables)
2. Where to get the data for the JPD (contains all possible combinations)?
3. Computation of sums:

$$\sum_{\boldsymbol{q'},\boldsymbol{h'}} P(\boldsymbol{q'}, \boldsymbol{e}, \boldsymbol{h'})$$

Number of **nested loops** = number of query variables + number of hidden variables

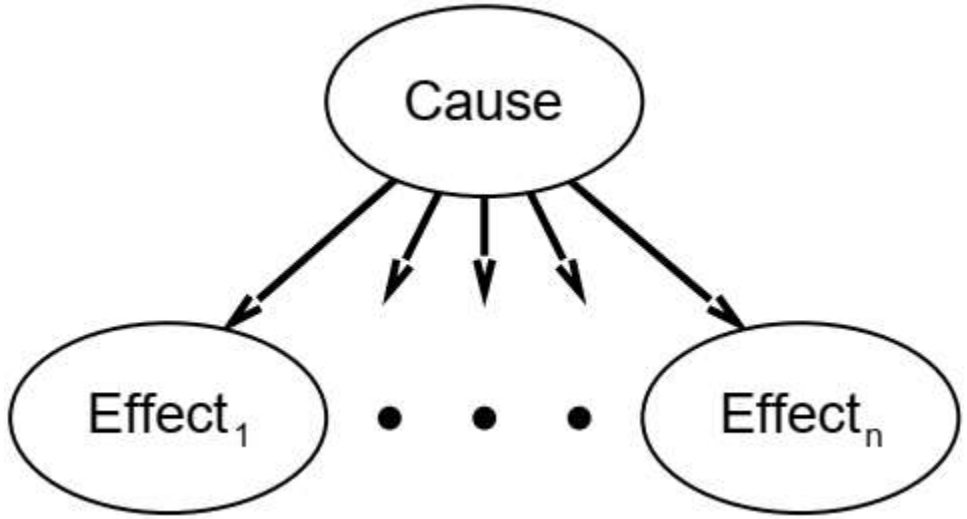
(again, exponential)

Computational Issues

Probabilistic inference is **NP hard**

In this course: simplified models

- approximate the JPD
- approximate computation



Example of a simplified model: Naïve Bayes

Independence

Decomposing Full Joint Distributions

Extreme Independence

X_j	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
										
$P(X_j = h)$	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5

Full joint distribution $P(x_1, \dots, x_{10})$

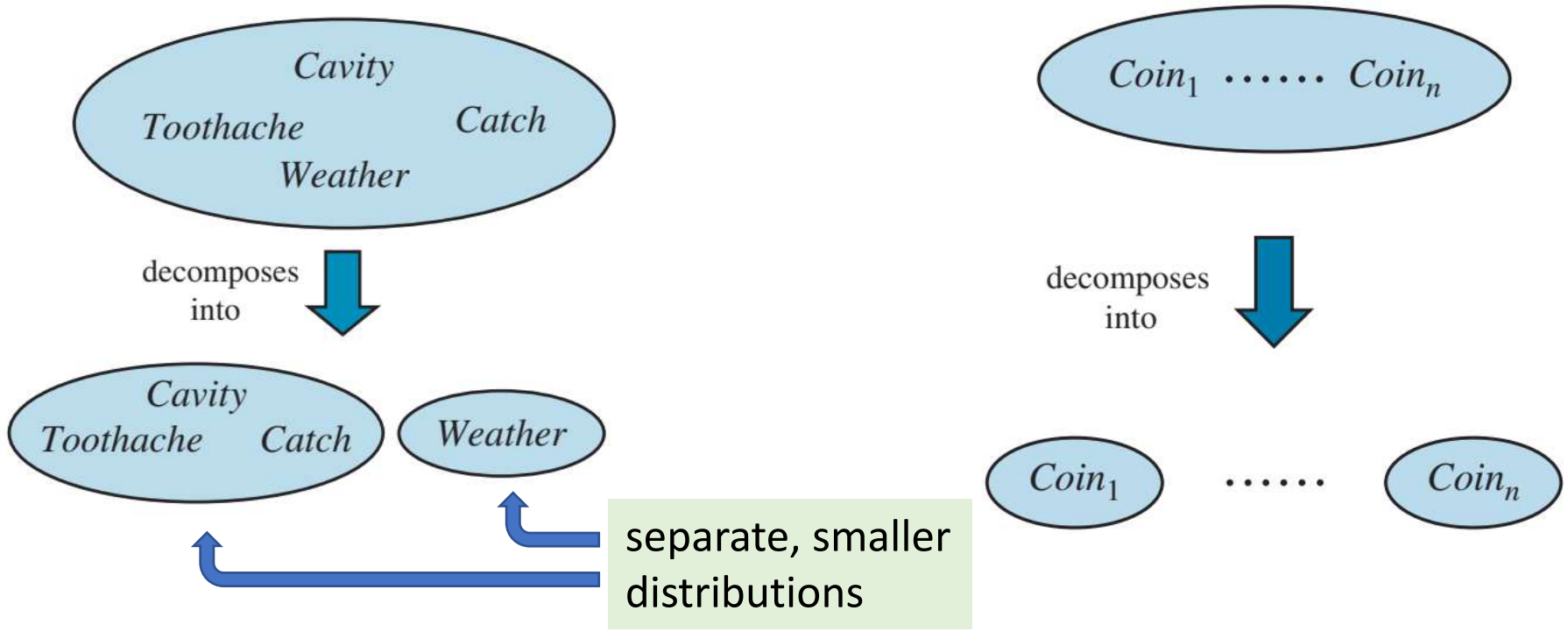
Each coin throw is independent, so

$$P(x_1, \dots, x_{10}) = P(x_1)P(x_2) \dots P(x_{10})$$

decomposes to product of individual P()-s

Decomposition

Examples of decomposition (because of independence)



Factorization

Factorization – partitioning the JPD into product of independent parts

Example: Weather is independent from dentistry

$$P(Weather, Cavity, Tootache, Catch)$$
$$= P(Weather)P(Cavity, Toothache, Catch)$$

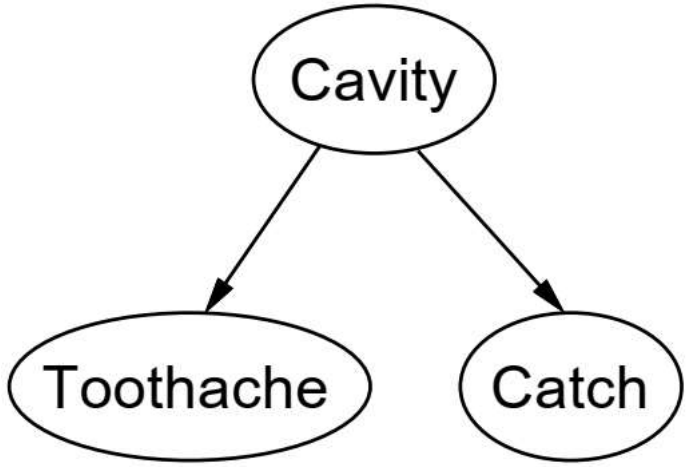
If sub-parts are small enough, complexity becomes near-linear

(coin example is truly linear)

Conditional Independence

$Toothache, Catch$ are not always independent:

$Cavity$ **not known**	$Cavity$ **known**
observing toothache indicates dental problems, so "catch" event becomes more likely	toothache doesn't cause the "catch" event. So has no influence on its probability.
NOT independent	independent



$P(A|B,C) = P(A|C)$ A is **conditionally independent** of B, given C

Conditional Independence

$P(A|B,C) = P(A|C)$ A is conditionally independent of B, given C

we can derive:

$$P(A,B|C) = P(A|C)P(B|C)$$

Only applies **if** A, B are conditionally independent! Note these can be sets, so for example:

$$P(x_1, x_2, x_3|C) = P(x_1|C)P(x_2|C)P(x_3|C)$$

we will use this one later

Naïve Bayes

A very simple but powerful model

Bayes' Rule

Derivation:

Take $P(A|B) = \frac{P(A,B)}{P(B)}$ (definition of cond. probability)

Rewrite $P(A,B) = P(B,A) = P(B|A)P(A)$

Result:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes' Rule

General form

$$P(\boldsymbol{q}|\boldsymbol{e}) = \frac{P(\boldsymbol{e}|\boldsymbol{q})P(\boldsymbol{q})}{P(\boldsymbol{e})}$$

Can be ignored in some cases

These probabilities are relatively easiest to estimate

- calculate for query variables $\boldsymbol{q}$, based on evidence $\boldsymbol{e}$
- Based on:
  - Query prior $P(\boldsymbol{q})$ (what the probability would be, without evidence)
  - $P(\boldsymbol{e}|\boldsymbol{q})$, usually some **causal relation**

Naïve Bayes Classifier

Let's rename things using machine learning terminology

Class C, features $\boldsymbol{x}$

$$P(C|\boldsymbol{x}) = \frac{P(\boldsymbol{x}|C)P(C)}{P(\boldsymbol{x})}$$

Now **assume** features are conditionally independent

(all are effects, caused directly by the class)

$$P(C|x_1, x_2, \dots, x_n) = \frac{P(C)P(x_1|C)P(x_2|C)\dots P(x_n|C)}{P(\boldsymbol{x})}$$

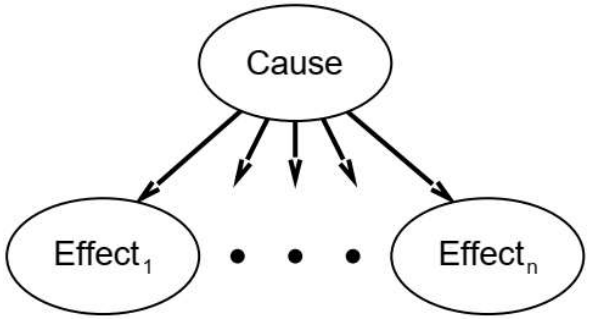
probability of seeing a feature, given some value for class, usually easy to learn!

we will get rid of this soon

Naïve Bayes Model

Says that variables have a structure:

(hidden) **cause** and it's (visible) **effects**



Model	cause	effects
Iris flowers	species	length/width of petals etc
Spam e-mails	intent of letter	choice of words in e-mail
Risk of Alzheimer's	genetic susceptibility	changes in G,C,A,T nucleic acids in specific genome locations

Computational Example

Data: https://github.com/sjwhitworth/golearn/blob/master/examples/datasets/tennis.csv

Goal: use weather forecast (“outlook”) and current observations (“temp”, “humidity”, “wind”)

Predict if the weather is suitable for playing tennis.

Naïve Bayes Classifier

Using the model to classify:

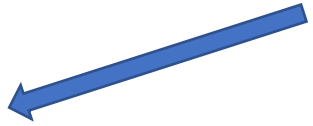
let's take $\alpha = \frac{1}{P(\boldsymbol{x})}$

$$P(C|x_1, x_2, \dots, x_n) = \alpha P(C)P(x_1|C)P(x_2|C) \dots P(x_n|C)$$

Compute $h(c_i) = P(c_i)P(x_1|c_i)P(x_2|c_i) \dots P(x_n|c_i)$

for each possible class c_i.

called
MAP hypothesis



The class with highest $h(c_i)$ is our prediction.

(α is the same each time so max $h(c_i)$ also gives max probability)

Prior Probability

Compute for each possible class c_i

$$h(c_i) = P(c_i)P(x_1|c_i)P(x_2|c_i) \dots P(x_n|c_i)$$

Find the probabilities by statistical estimate:

14 cases, 9 times the decision was “yes”

c_i	$P(c_i)$
no	$\frac{5}{14} = 0.36$
yes	$\frac{9}{14} = 0.64$

outlook	temp	humidity	windy	play
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
overcast	hot	high	FALSE	yes
rainy	mild	high	FALSE	yes
rainy	cool	normal	FALSE	yes
rainy	cool	normal	TRUE	no
overcast	cool	normal	TRUE	yes
sunny	mild	high	FALSE	no
sunny	cool	normal	FALSE	yes
rainy	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no

Conditional Probability for Observation

Compute for each possible class c_i

$$h(c_i) = P(c_i)P(x_1|c_i)P(x_2|c_i) \dots P(x_n|c_i)$$

Example: $P(x_1|"yes")$

x_1	$P(x_1\|"yes")$
overcast	$\frac{4}{9} = 0.44$
rainy	$\frac{3}{9} = 0.33$
sunny	$\frac{2}{9} = 0.22$

outlook	temp	humidity	windy	play
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
overcast	hot	high	FALSE	yes
rainy	mild	high	FALSE	yes
rainy	cool	normal	FALSE	yes
rainy	cool	normal	TRUE	no
overcast	cool	normal	TRUE	yes
sunny	mild	high	FALSE	no
sunny	cool	normal	FALSE	yes
rainy	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no

Conditional Probability for Observation

Compute for each possible class c_i

$$h(c_i) = P(c_i)P(x_1|c_i)P(x_2|c_i) \dots P(x_n|c_i)$$

Example: $P(x_1|"no")$

x_1	$P(x_1\|"no")$
overcast	$\frac{0}{5} = 0$
rainy	$\frac{2}{5} = 0.4$
sunny	$\frac{3}{5} = 0.6$

"Impossible" case in our model, this will mess up the calculation! Zero is caused by **not enough data**

outlook	temp	humidity	windy	play
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
overcast	hot	high	FALSE	yes
rainy	mild	high	FALSE	yes
rainy	cool	normal	FALSE	yes
rainy	cool	normal	TRUE	no
overcast	cool	normal	TRUE	yes
sunny	mild	high	FALSE	no
sunny	cool	normal	FALSE	yes
rainy	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no

Add-one Smoothing

Compute for each possible class c_i

$$h(c_i) = P(c_i)P(x_1|c_i)P(x_2|c_i) \dots P(x_n|c_i)$$

$$P(x_j|c_i) = \frac{n+1}{N+d}$$

n – cases with value x_j

N – cases with class c_i

d – how many values

x_1	$P(x_1\|"no")$
overcast	$\frac{1}{8} = 0.125$
rainy	$\frac{3}{8} = 0.375$
sunny	$\frac{4}{8} = 0.5$

outlook	temp	humidity	windy	play
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
overcast	hot	high	FALSE	yes
rainy	mild	high	FALSE	yes
rainy	cool	normal	FALSE	yes
rainy	cool	normal	TRUE	no
overcast	cool	normal	TRUE	yes
sunny	mild	high	FALSE	no
sunny	cool	normal	FALSE	yes
rainy	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no